# Predicting progression to Alzheimer's disease from clinical and imaging data: a reproducible study

Jorge Samper-González[1], Ninon Burgos[1], Simona Bottani[1], Marie-Odile Habert[2,3,4], Theodoros Evgeniou[5], Stéphane Epelbaum[1,6], Olivier Colliot[1,6,7]

[1]ARAMIS Lab, ICM, Inserm U1127, CNRS UMR 7225, Sorbonne University, Inria, Paris, France
[2]Laboratoire d'Imagerie Biomédicale, Sorbonne Université, Inserm U 1146, CNRS UMR 7371, Paris, France
[3]AP-HP, Hôpital Pitié-Salpêtrière, Department of Nuclear Medicine, Paris, France
[4]Centre Acquisition et Traitement des Images (CATI, cati-neuroimaging.com)
[5]INSEAD, Bd de Constance, 77305 Fontainebleau, France
[6]AP-HP, Department of Neuroradiology, Pitié-Salpêtrière Hospital, Paris, France
[7]AP-HP, Department of Neurology, Pitié-Salpêtrière Hospital, Paris, France

## Introduction

Various machine learning approaches have been developed for predicting progression to Alzheimer's disease (AD) in patients with mild cognitive impairment (MCI) from MRI and PET data (see e.g (Arbabshirani et al, 2017, Falahati et al, 2014, Rathore et al, 2017)). Objective comparison of these approaches is nearly impossible because of differences at all steps, from data management to image processing and evaluation procedures. Moreover, with a few exceptions such as (Moradi et al, 2015), these papers rarely compare their results to that obtained with clinical/cognitive data only, a critical point to demonstrate the practical utility of neuroimaging in this context. We previously proposed a framework for the reproducible evaluation of ML algorithms for AD classification (Samper-González et al, 2018). This framework was applied to AD classification using unimodal neuroimaging data (T1 MRI and FDG PET). Here, we extend our previous work to the combination of multimodal clinical and neuroimaging data for predicting progression to AD among MCI patients.
All the code is publicly available at: https://gitlab.icm-institute.org/aramislab/AD-ML.

## Methods

The data used was from the ADNI database. We included the same group of subjects as in (Samper-González et al, 2018), except 3 subjects who had missing cognitive tests. The population is described in Figure 1.
The framework comprises a modular set of tools from the Clinica software platform (www.clinica.run) for: converting ADNI into BIDS format, feature extraction from T1 MRI and

FDG-PET data using tools from SPM, classification algorithms based on scikit-learn and rigorous evaluation procedures. All features were corrected for the effect of age.
First, we tested models using clinical/cognitive data alone, based on a random forest classifier (RF). A first model used only gender, education level, MMSE and CDR-SB tests. We then assessed the added value of RAVLT (a memory test) and ADAS-Cog. We finally added APOE genotype.
Secondly, we tested imaging data (T1 MRI and FDG PET) in isolation. In a first approach, a SVM model was trained to distinguish pMCI vs sMCI. In a second approach, it was trained to separate CN ABeta- (amyloid negative cognitively normal subjects) and AD ABeta+ (amyloid positive AD patients) and applied to pMCI vs MCI classification.
Finally, we integrated clinical/cognitive and imaging data. For imaging data, we used as feature the continuous output of the SVM, which was combined with clinical/cognitive features into a RF classifier.

# Results

Results obtained with the different approaches are presented in Figure 2. The simplest clinical model performed poorly with a balanced accuracy (BAcc) of 0.66 and an area under the ROC curve (AUC) of 0.73. Inclusion of either RAVLT or ADAS-Cog led to a strong improvement (Bacc=0.75, AUC=0.84). Addition of APOE genotype did not improve the results. When trained on pMCI vs sMCI, imaging data performed poorly (BAcc=0.67 for T1w MRI, BAcc=0.71 for PET). Training on CN ABeta- vs AD ABeta+ substantially improved the results for FDG PET (BAcc=0.76, AUC=0.82), the performance being close to that of clinical models. Finally, combination of clinical and imaging data further improved the results, although moderately (BAcc=0.79, AUC= 0.89).

# Conclusions

We demonstrated that using only clinical data allows reaching acceptable performance levels. Integrating both imaging and clinical data can further improve the accuracy, but the improvement remains moderate. We further showed that a simple trick (training on the easier task CN ABeta- vs AD Beta+) improved imaged-based results. Moreover, using standard ML approaches, we reached performances that are comparable to published sophisticated algorithms (see e.g. (Rathore et al, 2017)).
Overall, our results highlight the importance of systematically comparing image-based ML approaches to a baseline using clinical data, an issue that is often overlooked in the neuroimaging community.

# References

(Arbabshirani et al, 2017) Arbabshirani, M.R., Plis, S., Sui, J., Calhoun, V.D., 2017. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. Neuroimage 145, 137–165. https://doi.org/10.1016/j.neuroimage.2016.02.079

(Falahati et al, 2014) Falahati, F., Westman, E., Simmons, A., 2014. Multivariate Data Analysis and Machine Learning in Alzheimer's Disease with a Focus on Structural Magnetic Resonance Imaging. Journal of Alzheimer's Disease 41, 685–708. https://doi.org/10.3233/JAD-131928

(Moradi et al, 2015) Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J., Alzheimer's Disease Neuroimaging Initiative, 2015. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. Neuroimage 104, 398–412. https://doi.org/10.1016/j.neuroimage.2014.10.002

(Rathore et al, 2017) Rathore S., Habes M., Iftikhar M.A., Shacklett A., Davatzikos C. (2017), 'A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages', Neuroimage, vol. 155, pp. 530-548

(Samper-González et al, 2018) Samper-González, J., Burgos, N., Bottani, S., Fontanella, S., Lu, P., Marcoux, A., Routier, A., Guillon, J., Bacci, M., Wen, J., Bertrand, A., Bertin, H., Habert, M.-O., Durrleman, S., Evgeniou, T., Colliot, O., Alzheimer's Disease Neuroimaging Initiative, Australian Imaging Biomarkers and Lifestyle flagship study of ageing, 2018. Reproducible evaluation of classification methods in Alzheimer's disease: Framework and application to MRI and PET data. 183, pp 504-521. Neuroimage. https://doi.org/10.1016/j.neuroimage.2018.08.042

# Figures

Figure 1. Studied populations. Summary of participant demographics, mini-mental state examination (MMSE) and global clinical dementia rating (CDR) scores

| | **N** | **Age*** | **Gender** | **MMSE*** | **CDR** |
|---|---|---|---|---|---|
| $sMCI_{36}$ | 340 | 71.8±7.5 [55.0, 88.6] | 201 M / 139 F | 28.1±1.6 [23, 30] | 0.5: 340 |
| $pMCI_{36}$ | 167 | 74.9±6.9 [55.0, 88.3] | 98 M / 69 F | 27.0±1.7 [24, 30] | 0.5: 166; 1: 1 |
| CN $A\beta$- | 115 | 72.2±6.1 [56.2, 89.0] | 59 M / 56 F | 29.0±1.3 [24,30] | 0: 115 |
| AD $A\beta$+ | 126 | 74.1±8.1 [55.1, 90.3] | 65 M / 61 F | 22.9±2.1 [19, 26] | 0.5: 54; 1: 71; 2: 16 |

* Values are presented as mean±SD [range]. M: male, F: female

Figure 2. Results for models using clinical/cognitive data only (first block), imaging data only (second block) and combination of imaging and clinical/cognitive data (third block).
$\text{Clinical}_{\text{base}}$: gender, education level, MMSE score, sum of boxes of CDR test. RAVLT: Rey Auditory Verbal Learning Test. ADAS-Cog: scores of Alzheimer's Disease Assessment Scale cognitive sub-scale separated into four categories (memory, language, concentration and praxis. APOE4: number of E4 alleles for APOE (apolipoprotein E) genotype. T1w MRI: T1-weighted magnetic resonance imaging. FDG-PET: fluorodeoxyglucose positron emission tomography. $\text{Score}_{\text{T1}}$ (resp. $\text{Score}_{\text{FDG}}$): continuous output of the SVM using T1w MRI (resp. FDG-PET) data.

| **Classifier - Features** | **Bal. acc.** | **AUC** | **Acc.** | **Sens.** | **Spec.** |
|---|---|---|---|---|---|
| RF - $\text{Clinical}_{\text{base}}$ | 0.660 | 0.726 | 0.684 | 0.587 | 0.734 |
| RF - $\text{Clinical}_{\text{base}}$ + RAVLT | **0.742** | 0.823 | 0.75 | 0.717 | 0.767 |
| RF - $\text{Clinical}_{\text{base}}$ + ADAS | **0.754** | 0.836 | 0.760 | 0.736 | 0.772 |
| RF - $\text{Clinical}_{\text{base}}$ + RAVLT + ADAS | **0.762** | 0.852 | 0.768 | 0.743 | 0.781 |
| RF - $\text{Clinical}_{\text{base}}$ + RAVLT + APOE4 | 0.756 | 0.838 | 0.759 | 0.750 | 0.763 |
| RF - $\text{Clinical}_{\text{base}}$ + ADAS + APOE4 | 0.757 | 0.842 | 0.766 | 0.731 | 0.784 |
| RF - $\text{Clinical}_{\text{base}}$ + RAVLT + ADAS + APOE4 | 0.765 | 0.857 | 0.772 | 0.746 | 0.785 |
| SVM - T1w MRI | 0.670 | 0.736 | 0.698 | 0.586 | 0.754 |
| SVM (trained on CN $A\beta$- vs AD $A\beta$+) - T1w MRI | 0.679 | 0.764 | 0.708 | 0.547 | 0.811 |
| SVM - FDG PET | 0.708 | 0.777 | 0.732 | 0.633 | 0.782 |
| SVM (trained on CN $A\beta$- vs AD $A\beta$+) - FDG PET | **0.761** | 0.818 | 0.788 | 0.666 | 0.856 |
| RF - $\text{Clinical}_{\text{base}}$ + $\text{Score}_{\text{T1}}$ | 0.717 | 0.792 | 0.732 | 0.671 | 0.763 |
| RF - $\text{Clinical}_{\text{base}}$ + $\text{Score}_{\text{FDG}}$ | 0.760 | 0.831 | 0.791 | 0.669 | 0.852 |
| RF - $\text{Clinical}_{\text{base}}$ + $\text{Scores}_{\text{T1,FDG}}$ | 0.769 | 0.855 | 0.796 | 0.685 | 0.852 |
| RF - $\text{Clinical}_{\text{base}}$ + RAVLT + $\text{Scores}_{\text{T1,FDG}}$ | **0.791** | 0.881 | 0.809 | 0.735 | 0.846 |
| RF - $\text{Clinical}_{\text{base}}$ + ADAS + $\text{Scores}_{\text{T1,FDG}}$ | **0.790** | 0.873 | 0.810 | 0.729 | 0.851 |
| RF - $\text{Clinical}_{\text{base}}$ + RAVLT + ADAS + $\text{Scores}_{\text{T1,FDG}}$ | **0.792** | 0.888 | 0.811 | 0.736 | 0.849 |